

Meta data category and a method of building an information portal

The invention relates to the Internet (also known as the World Wide Web WWW). In particular, the invention relates to the provision of a new meta data category and a method of building an information portal and personalizing that portal in accordance with a user profile and utilizing the new meta data category.

5

The Internet is a large network of computers in which a user may access information from one of any number of Web servers over the network, such Web servers being physically located anywhere in the world. Essentially, the user makes a request over the Internet and the relevant Web server passes the requested information to the user. The information available on such Web servers may comprise text, graphics, video or audio. In addition, Web servers may provide access to user executable applications. Because of the vast amount of information stored on Web servers and accessible via the Internet, users can employ search engines to find information which is relevant to their particular needs. In order for a search engine to carry out this particular function the information needs to be indexed or stored in database structures.

10

15

Web portals offer search engine functionality and also access to a predefined set of applications available locally or distributed on Web servers across the globe. There are currently no means available to search the behaviour or functionality of applications presented by a portal. As the number of applications available is growing rapidly it will become increasingly difficult to find the right application or service on the Internet in the future. This problem is referred to as "application overload".

20

25

It is a first aim of embodiments of the present invention to provide a means for identifying applications or services relevant to a particular user.

It is another aim of embodiments of the present invention to provide a method for generating a portal to information and applications on the Internet based on a personal profile and/or feedback from user queries to reduce application overload.

According to a first aspect of the invention, there is provided a meta data category specifying types of web applications and services.

Such provision brings a new versatility to the web and is a major help in overcoming the application overload problem.

5 Preferably, within said meta data category a plurality of functional identifiers may be present, each functional identifier denoting a specific type of application or service available via the world wide web.

The functional identifiers are building blocks for a task description language for describing the functionality of a web application or service. A web application document
10 would contain a piece of text (or a link to a separate document) written in such a task description language.

Providing a functional identifier for web based applications and services supplements existing content identifiers and thereby helps users locate specific services/applications.

15 Preferably, each functional identifier comprises one or more keywords identifying a type of application or service being offered by a web based provider.

Key word identifiers enable efficient classification by search engines.

Preferably, functional identifiers are written in a task description language which describes the functionality of the application or service.

20 Preferably, the meta data category is provided in a searchable area of a web page, preferably in a header region. Alternatively, it may be provided as part of a URL (Uniform Resource Locator).

Providing the meta data category in the meta header or other searchable region enables functional identifiers to be read by dedicated web crawlers when building up search
25 terms for use with search engines etc.

Typical applications or services identified by such functional identifiers may, for instance, comprise manufacturing, sales or repairing of physical items, the provision of professional, domestic or other services, etc.

According to a second aspect of the invention, there is provided a method of
30 building up a portal to information and applications on a network comprising retrieving content from one or more data sources that match the user profile and presenting the retrieved content to a user's web browser, wherein the retrieved content presented to the user includes information concerning web applications and/or services, said information being retrieved on the basis of web application and/or service types.

Providing such a portal enables for the first time an effective transmission of information to users which identifies services relevant to a user's needs or interests.

Preferably, the method further comprises analysing an input user query to detect a desire for retrieving web applications and/or services, retrieving content from one or more data sources that match the user query and presenting the retrieved content to the users web browser.

Preferably, displayed results presented to the user via the user's web browser comprises conventional content as well as information concerning web applications/services and such conventional content and the information are merged prior to presentation to the user - merging may be accomplished in a generation module. Retrieved applications may be regarded as functional content.

The step(s) of retrieving content from data sources may comprise filtering content and applications from said one or more data sources to match such content with a user query or user profile.

Preferably, the web browser is configured to deliver information to the user, based on a user profile, to personalise the portal.

Personalisation does not have to happen inside the browser, the presentation in the form of a web document containing the personalised portal interface is preferably generated on the server. However, the user profile may be used to personalise the look-and-feel of the browser.

Preferably, the user profile is constructed based upon implicit feedback from information originating from the user and delivered by the browser. Such implicit feedback may comprise determining patterns of user behaviour from user input to the browser so as to build up and modify knowledge about the user and store such knowledge in the user profile. Such knowledge may also comprise knowledge about domain, platform and general presentation styles etc.

In the step of retrieving content, content may be retrieved based directly upon the user query to retrieve applications.

In the step of retrieving content, said content is preferably retrieved based directly upon a user query and such content includes one or more links to application and/or service providers identified by said one or more functional identifiers.

Preferably, the meta data category is a meta data category in accordance with the first aspect of the invention.

For a better understanding of the invention, and to show how embodiments for the same may be carried into effect, reference will now be made by way of example, to the accompanying diagrammatic drawings in which:

Fig. 1 is a schematic diagram illustrating the indexing and retrieval of Web applications and services using meta-data described in a task-description language; and

Fig. 2 shows a schematic diagram illustrating an architecture for information processing on the Web according to an aspect of the present invention.

As mentioned in the discussion of the prior art, meta data categories currently employed tend to designate content of web pages or types of content present on web pages (e.g. whether video is present, audio etc.). To address the problem of application overload, the inventors propose a new meta data category which provides functional indications according to applications or services being offered via a web site. Such functional indications can thereby provide a task-based indexing of web services, which is extremely valuable for information portals. An example of such an information portal may be found at the URLs (Uniform Resource Locators) www.yahoo.com, www.altavista.com etc. It is possible to personalise such portals to an extent so as to tailor a general web portal into something which is a little more specific (e.g. My Yahoo, My Excite).

However, personalisation in terms of retrieving information concerning functionality and services offered by the web is at present impractical.

To generate a portal in accordance with an aim of the present invention requires the existence of an appropriate meta-data category which designates "functionality".

With the use of meta-data, Web applications that show complex behavior are regarded as objects described by meta-data, and the descriptions stored in database tables to be retrieved later. This can be accomplished if textual descriptions are added to existing uniform resource locators (URLs) or a specific field is added to a header region of a web page.

There is shown below a typical header region written in HTML (Hyper Text Markup Language) of a conventional web site:

<HTML>

<HEAD>

<META name="description" content="xxxxxxxxxx">

<TITLE>xxxxxx</TITLE>
</HEAD>

In the above example, there is shown a content oriented meta data listing provided in an HTML header region of a web page. In the newly proposed system, functional, rather than content type meta data is included in the header region.

Such meta data may alternatively be included in any other area of the web page, such as in the body region enclosed in special XML tags - so long as it does not affect the presentation of the web document in a browser and can be understood by machines (crawlers, indexers).

In a preferred embodiment of the present invention a functionally oriented meta-data category describing Web applications is written in a task-description language (TDL) that describes the functionality of the Web application or service as one or more functional identifiers. These functional identifiers can be indexed or stored into database tables according to the functions or types of applications that they represent and be retrieved by the user, as shown in Fig. 1.

Fig. 1 is a schematic diagram illustrating the indexing and retrieval of Web applications and services using meta-data described in a task-description language in which there is shown a browser 10 and a database 12 built up of task-description language (TDL) documents 14.

The user inputs a task query to their browser, which in accordance with their personal profile and the current query searches the task-description language (TDL) documents indexed in a database, and presents the user with relevant applications. The browser can then update the user profile based on knowledge of the input query and/or the applications selected by the user from the list presented.

The user input could also be converted into TDL by a user input query formulation unit (using a Graphical User Interface for example) if the TDL language is too complex for ordinary users.

A suitable format for the meta-data is the Resource Description Framework (RDF) which is under development by the World Wide Web consortium (W3C) (see www.w3.org/RDF/), but any similar and widely acceptable language could be used to record and store the necessary meta-data. The database 12 of TDL documents 14 is compiled by a Web crawler that looks for new or modified content on the Web to index. Web applications

and services must be represented in a machine understandable way for the Web crawler to correctly index them, and this functionality is provided by the same meta-data.

Figure 2 shows a high level overview of the architecture of the new semi-automatic personalised portal generation and presentation process. The architecture consists

of three layers; a knowledge layer KL that consists of knowledge bases on the Web, a document layer DL that represents, retrieves and presents information on the Web using the knowledge layer KL and a portal layer PL which delivers personalised Web applications and services to different users. The portal layer PL comprises an application indexing block 201 and an application retrieval block 202. The document layer DL comprises a document discovery block 211, a document indexing block 212, a database management system and/or information retrieval block 213 (DBMS/IR), a presentation generation block 214, a user input analysis block 215 and a Web browser 216. The knowledge layer KL comprises a knowledge discovery block 221, a knowledge extraction block 222, a knowledge reasoning block 223, a knowledge modification block 224, knowledge bases 225 and a locally stored user profile block 226.

The Web browser 216 allows a user to interact with Web content. In most cases, the Web browser 216 will be a software component that displays Web content, although it does not necessarily have to display such content on a screen, as speech dialogue systems and Braille systems also allow users to interact with Web content through specifically designed interfaces. The Web browser 216 supplies implicit feedback information about the user to the user input analysis block 215 which identifies patterns in the users behaviour that can be used to modify knowledge about the user. Also, the browser 216 may pass other information to the user input analysis block 215 regarding the platform on which it is running, the users domain and general presentation styles preferred by the user etc. The Web browser 216 may in addition be adapted to allow the user to provide explicit feedback about the type of application and/or information required. The user input analysis block 215 controls the creation and maintenance of a personalised view on the Web that best matches the users requirements at a given point in time.

The user input analysis block 215 then formulates a query and passes it to the DBMS/IR block 213 which determines the content or applications that best match that query and feeds this information to the presentation generation block 214.

The presentation generation block 214 receives information from the DBMS/IR block 213 and the application retrieval block 202 and filters this information to create a portal which gives a coherent view on information that matches the users information

needs. The presentation generation block 214 can then supply a personalised set of hyper-linked documents, which can comprise plain text documents, multimedia objects or Web applications to the Web browser 216 for communication with the user.

At no point in the procedure is the user concerned by details of where information is gathered from. The document discovery block 211 is a Web crawler, which represents documents and applications on the Web in a machine understandable way and passes reference to them to the document indexing block 212 and application indexing block 201. Text documents may be indexed using keywords, but the use of meta-data is required to provide suitable identification for applications and thereby allow them to be indexed.

The ability to adapt to a users personal needs by incorporating knowledge gained implicitly over time is useful to some degree for all the components shown in Fig. 2. For example, the document discovery block 211 might be personalised to search for information on applications in a specific domain. However, before knowledge can be used to adapt these processes, it must first be extracted automatically, or modelled manually. The knowledge discovery block 221 locates meta-data on the Web that can be extracted or modelled by the knowledge extraction block 222, this knowledge is then placed in the knowledge bases 225. The knowledge bases 225 can either store information locally, or on the Web depending on the type of knowledge, the amount of knowledge, the performance required in accessing the knowledge, the privacy of the knowledge or other factors.

Besides direct acquisition of knowledge from documents or profiles on the Web, knowledge can also be obtained and updated through implicit, explicit or social input from the users which is analysed by the user input analysis block 215 and then passed to the knowledge modification block 224 for storage in the knowledge bases 225. This process ensures that the knowledge stored in the knowledge bases 225 is up to date, and therefore useful to when passed to the other components shown in Fig. 2. The knowledge reasoning block 223 retrieves knowledge from the knowledge bases 225 and distributes it to the document discovery block 211, the document indexing block 212, the DBMS/IR block 213, the application indexing block 201, the presentation generation block 214, and the user input analysis block 215 which can then all use this knowledge to adapt accordingly.

From the above, it will be evident to the man skilled in the art that the present invention provides an effective form of categorising applications or services which are offered by the web according to *functionality*. Also, based on the newly proposed meta data category designating such functionality there is described a method by which information

2711 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2